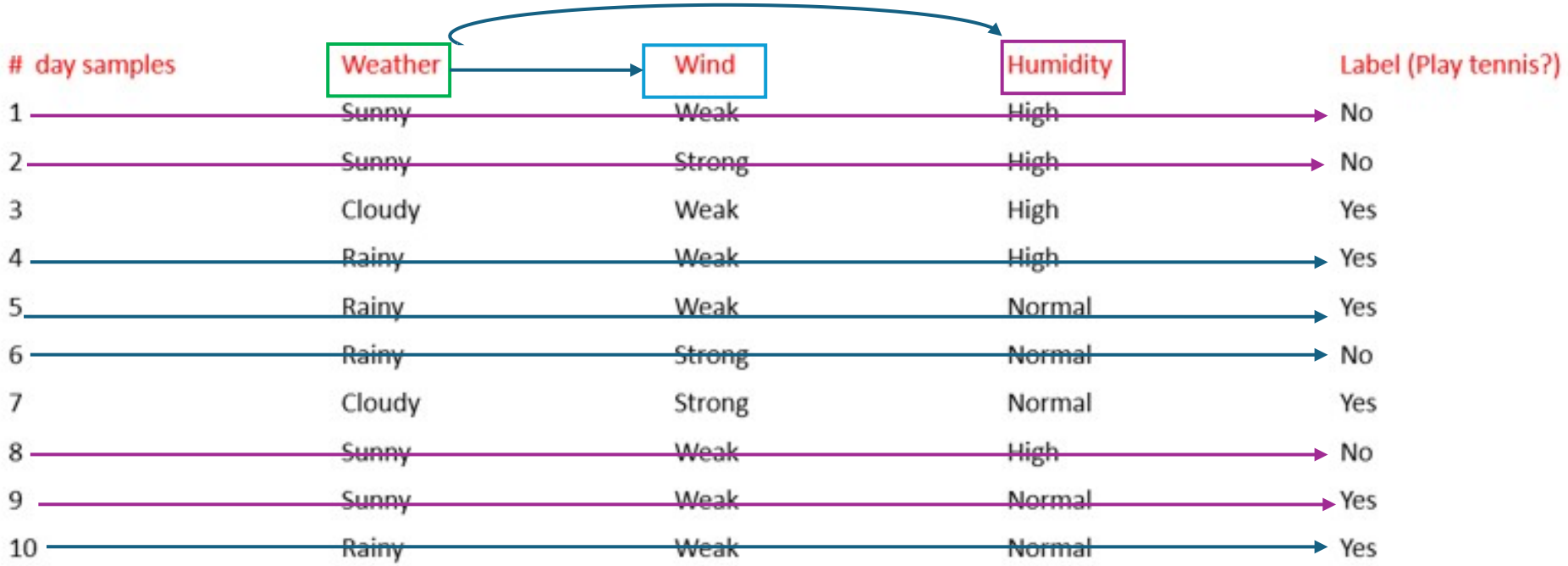
Bayesian Networks

Luis G. Jaimes

Intro (are the features independent ?)

- We already classified this dataset with Naive Bayes. The model worked, but it assumed that Weather, Wind, and Humidity were completely independent given the label. Let's ask whether that assumption is reasonable.
- Think about it intuitively:
- On rainy days, is wind more likely to be strong or weak? 3 out 4
- On sunny days, does humidity tend to be high or low?

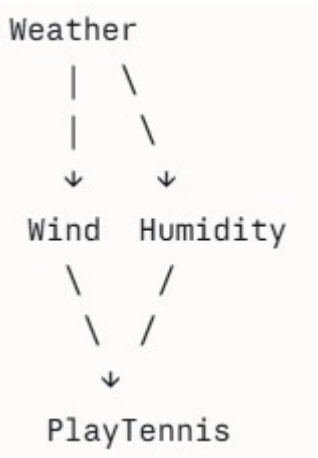


# day samples	Weather	Wind	Humidity	Label (Play tennis?)
1	Sunny	Weak	High	No
2	Sunny	Strong	High	No
3	Cloudy	Weak	High	Yes
4	Rainy	Weak	High	Yes
5	Rainy	Weak	Normal	Yes
6	Rainy	Strong	Normal	No
7	Cloudy	Strong	Normal	Yes
8	Sunny	Weak	High	No
9	Sunny	Weak	Normal	Yes
10	Rainy	Weak	Normal	Yes

Example

Wind and Humidity are also atmospheric; they are influenced by Weather patterns. The decision to play tennis is then a consequence of all three conditions.

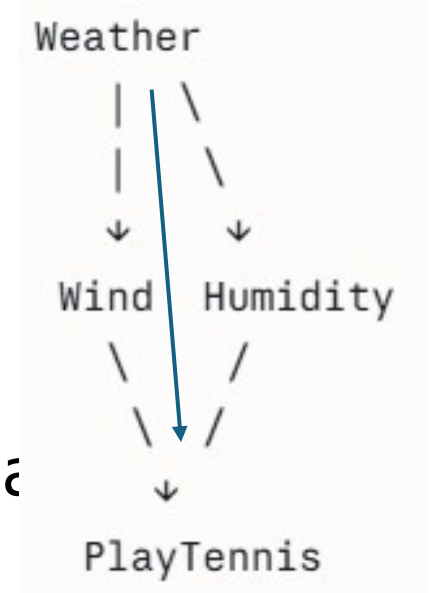
# day samples	Weather	Wind	Humidity	Label (Play tennis?)
1	Sunny	Weak	High	No
2	Sunny	Strong	High	No
3	Cloudy	Weak	High	Yes
4	Rainy	Weak	High	Yes
5	Rainy	Weak	Normal	Yes
6	Rainy	Strong	Normal	No
7	Cloudy	Strong	Normal	Yes
8	Sunny	Weak	High	No
9	Sunny	Weak	Normal	Yes
10	Rainy	Weak	Normal	Yes



Reading this graph:

- Weather directly influences Wind and Humidity
- Weather, Wind, and Humidity all directly influence PlayTennis
- Wind and Humidity do not influence each other directly, any correlation between them is explained by their shared parent Weather

Steps

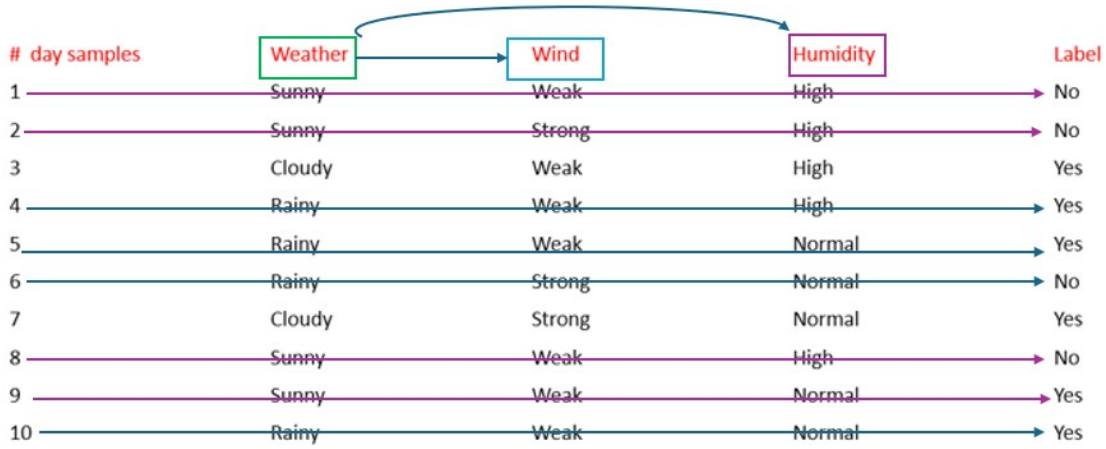


- Step 1: Think about the structure before touching the da
- Step 2: Get the priors for root nodes (CPT for Weather)
- Step 3: Build CPT for Wind given Weather P(wind | weather)
- Step 4: Build CPT for Humidity given Weather
- Step 5: Build CPT for PlayTennis given Weather, Wind, Humidity

Steps

Step 2: Priors for the root, CPTs for weather

Weather	P
Sunny	4/10 = 0.40
Cloudy	2/10 = 0.20
Rainy	4/10 = 0.40



# day samples	Weather	Wind	Humidity	Label
1	Sunny	Weak	High	No
2	Sunny	Strong	High	No
3	Cloudy	Weak	High	Yes
4	Rainy	Weak	High	Yes
5	Rainy	Weak	Normal	Yes
6	Rainy	Strong	Normal	No
7	Cloudy	Strong	Normal	Yes
8	Sunny	Weak	High	No
9	Sunny	Weak	Normal	Yes
10	Rainy	Weak	Normal	Yes

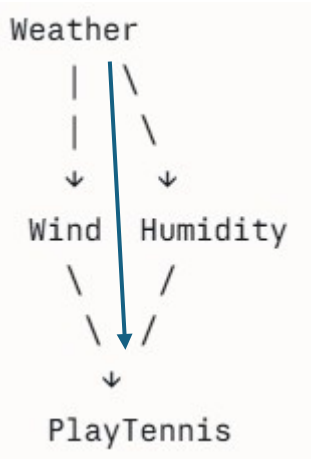
Step 3: CPT for wind given weather P(wind | weather)

Weather	P(Weak)	P(Strong)
Sunny	3/4 = 0.75	1/4 = 0.25
Cloudy	1/2 = 0.50	1/2 = 0.50
Rainy	3/4 = 0.75	1/4 = 0.25

When Weather = Sunny (rows 1, 2, 8, 9)
Then, how many out of the sunny are wind = weak (3)

Step4: CPT for wind given weather P(humidity | weather)

Weather	P(High)	P(Normal)
Sunny	3/4 = 0.75	1/4 = 0.25
Cloudy	1/2 = 0.50	1/2 = 0.50
Rainy	1/4 = 0.25	3/4 = 0.75



Step: P(playtennis | weather, wind, humidity)

Step 5:

Go through every unique combination that appears:

- **Sunny, Weak, High** → rows 1, 8 → both No → P(Yes)=0, P(No)=1
- **Sunny, Strong, High** → row 2 → No → P(Yes)=0, P(No)=1
- **Sunny, Weak, Normal** → row 9 → Yes → P(Yes)=1, P(No)=0
- **Cloudy, Weak, High** → row 3 → Yes → P(Yes)=1, P(No)=0
- **Cloudy, Strong, Normal** → row 7 → Yes → P(Yes)=1, P(No)=0
- **Rainy, Weak, High** → row 4 → Yes → P(Yes)=1, P(No)=0

P(PlayTennis = yes | weather = sunny , wind = weak , humidity = high) □ rows 1, 8 P(NO) =2/2, P(yes)=0

# day samples	Weather	Wind	Humidity	Label (Play tennis?)
1	Sunny	Weak	High	No
2	Sunny	Strong	High	No
3	Cloudy	Weak	High	Yes
4	Rainy	Weak	High	Yes
5	Rainy	Weak	Normal	Yes
6	Rainy	Strong	Normal	No
7	Cloudy	Strong	Normal	Yes
8	Sunny	Weak	High	No
9	Sunny	Weak	Normal	Yes
10	Rainy	Weak	Normal	Yes

Steps

- Step 5: Build CPT for PlayTennis given Weather, Wind, Humidity

  P(PlayTennis | Weather, Wind, Humidity)

Weather	Wind	Humidity	P(Yes)	P(No)
Sunny	Weak	High		
Sunny	Strong	High		
Sunny	Weak	Normal		
Sunny	Strong	Normal		
Cloudy	Weak	High		
Cloudy	Strong	Normal		
Cloudy	Weak	Normal		
Cloudy	Strong	High		
Rainy	Weak	High		
Rainy	Weak	Normal		
Rainy	Strong	Normal		
Rainy	Strong	High		

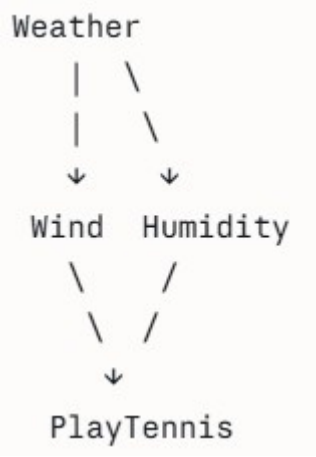
Steps

- Step 5: Build CPT for PlayTennis given Weather, Wind, Humidity

  P(PlayTennis | Weather, Wind, Humidity)

Weather	Wind	Humidity	P(Yes)	P(No)
Sunny	Weak	High	0/2 = 0.00	2/2 = 1.00
Sunny	Strong	High	0/1 = 0.00	1/1 = 1.00
Sunny	Weak	Normal	1/1 = 1.00	0/1 = 0.00
Sunny	Strong	Normal	—	—
Cloudy	Weak	High	1/1 = 1.00	0/1 = 0.00
Cloudy	Strong	Normal	1/1 = 1.00	0/1 = 0.00
Cloudy	Weak	Normal	—	—
Cloudy	Strong	High	—	—
Rainy	Weak	High	1/1 = 1.00	0/1 = 0.00
Rainy	Weak	Normal	2/2 = 1.00	0/2 = 0.00
Rainy	Strong	Normal	0/1 = 0.00	1/1 = 1.00
Rainy	Strong	High	—	—

Classify a new observation



Query: **Weather=Sunny, Wind=Weak, Humidity=Norma**

P(T | Sunny, Weak, Normal) = P(T, Sunny, Weak, Normal) / P(Sunny, Weak, Normal)

P(T, Sunny, Weak, Normal) = P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny) × P(T|Sunny, Weak, Normal)

The denominator is obtained by summing the numerator over all values of T:

P(Sunny, Weak, Normal) = P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny) × P(Yes|Sunny, Weak, Normal) +

P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny) × P(No|Sunny, Weak, Normal)

Factor out the common terms:

P(Sunny, Weak, Normal) = P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny) × [P(Yes|Sunny,Weak,Normal) + P(No|Sunny,Weak,Normal)]

Since P(Yes|Sunny,Weak,Normal) + P(No|Sunny,Weak,Normal) = 1:

P(Sunny, Weak, Normal) = P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny)

Query: **Weather=Sunny, Wind=Weak, Humidity=Normal ?**

P(T | Sunny, Weak, Normal) = P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny) × P(T|Sunny, Weak, Normal)/P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny)

The: P(Sunny) × P(Weak|Sunny) × P(Normal|Sunny) terms cancel and you are left with:

P(T | Sunny, Weak, Normal) = P(T | Sunny, Weak, Normal)

Weather	Wind	Humidity	P(Yes)	P(No)
Sunny	Weak	Normal	1/1 = 1.00	0/1 = 0.00

Getting the priors and conditional probs

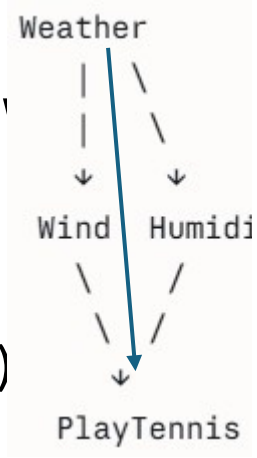
Build CPT for Play Tennis given Weather, Wind, Humidity P(play tennis | weather, wind, humidity)

P(W, Wi, H, T) = P(W) × P(Wi|W) × P(H|W) × P(T|

Compare the naïve bayes which assumes indepe

P(W, Wi, H, T) = P(T) × P(W|T) × P(Wi|T) × P(H|T)

$$P(C_k \mid x_1, x_2, \ldots, x_n) = \frac{P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)}{P(x_1, x_2, \ldots, x_n)}$$



Conditionally Probability Table (CPT) for Weather:

Weather	P
Sunny	4/10 = 0.40
Cloudy	2/10 = 0.20
Rainy	4/10 = 0.40

CPT for wind given weather P(wind | weather)

Weather	P(Weak)	P(Strong)
Sunny	3/4 = 0.75	1/4 = 0.25
Cloudy	1/2 = 0.50	1/2 = 0.50
Rainy	3/4 = 0.75	1/4 = 0.25

CPT for wind given weather P(humidity | weather)

Weather	P(High)	P(Normal)
Sunny	3/4 = 0.75	1/4 = 0.25
Cloudy	1/2 = 0.50	1/2 = 0.50
Rainy	1/4 = 0.25	3/4 = 0.75

P(A|B) = P(A,B)/P(B)
P(PlayTennis | Weather, Wind, Humidity) = P(PlayTennis, Weather, Wind, Humidity)/ P(Weather, Wind, Humidity)

Note: The denominator is the same thing summed over both values of Play Tennis (Normalizer)

P(Weather) × P(Wind|Weather) × P(Humidity|Weather) × P(PlayTennis|Weather, Wind, Humidity) /
(P(Weather) × P(Wind|Weather) × P(Humidity|Weather) × P(Yes|Weather, Wind, Humidity) + P(Weather) × P(Wind|Weather) × P(Humidity|Weather) × P(No|Weather, Wind, Humidity))

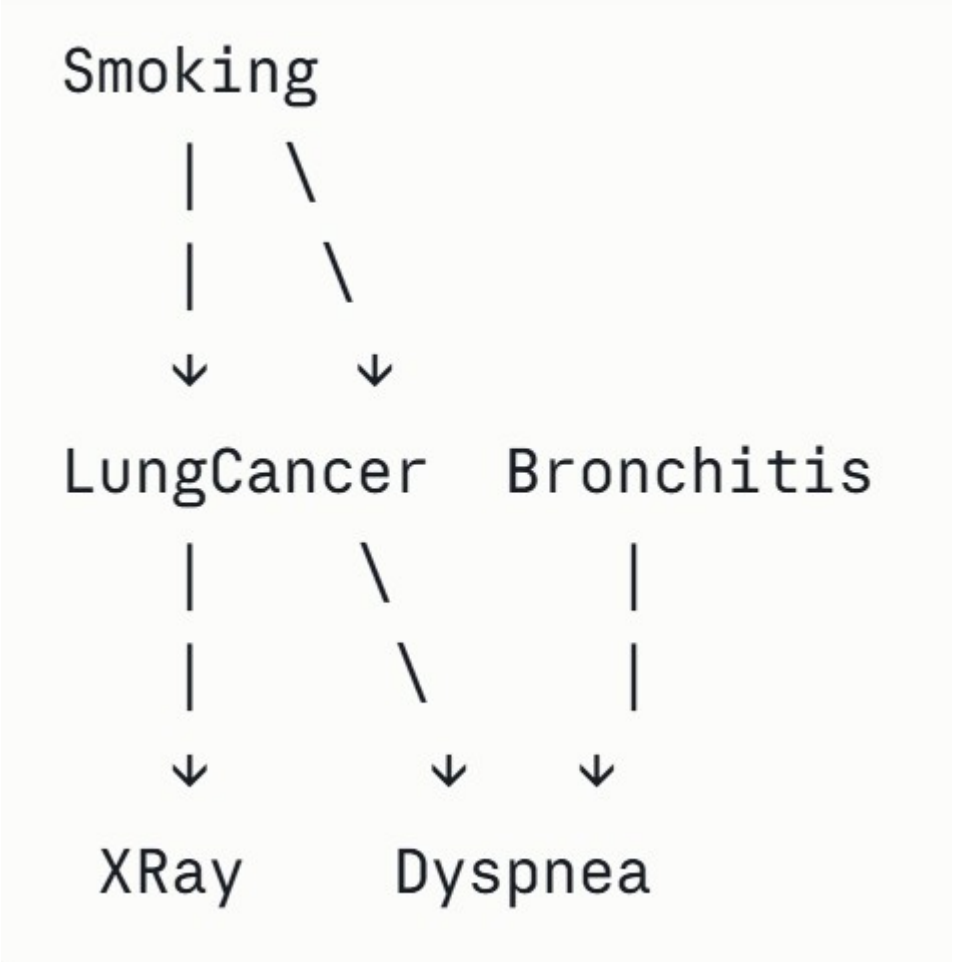
Weather = Rainy, Wind = Weak, Humidity = High,
play= ?

Example 2: A Medical Bayesian Network, Diagnosing Lung Disease

A doctor wants to diagnose whether a patient has lung cancer or bronchitis based on symptoms and risk factors

Variable	Values
`Smoking`	Yes, No
`LungCancer`	Yes, No
`Bronchitis`	Yes, No
`Dyspnea` (shortness of breath)	Yes, No
XRay (chest X-ray result)	Positive, Negative

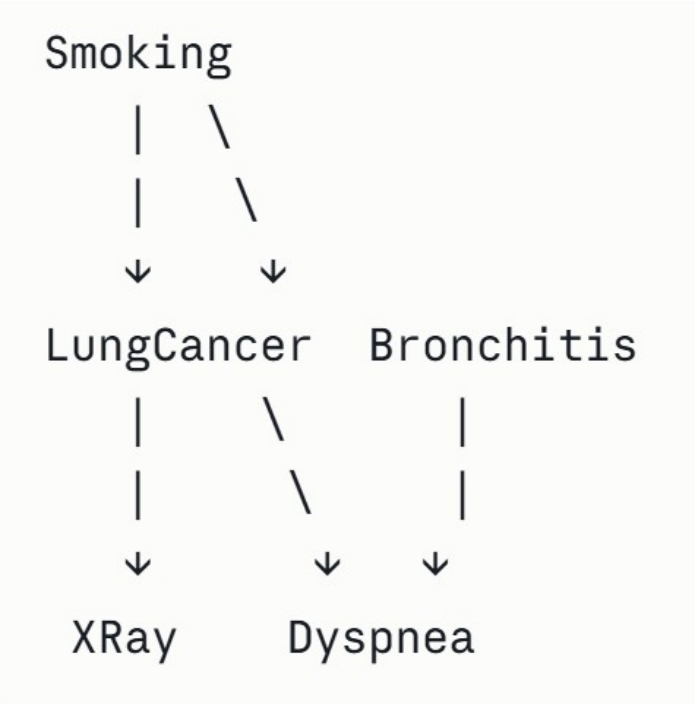
Example: causality



- Smoking directly causes both LungCancer and Bronchitis
- LungCancer causes a positive XRay result
- Both LungCancer and Bronchitis cause Dyspnea
- Smoking is a root node, no parents, just a prior

Example 2: Find the join factorization

P(S, L, B, X, D) =

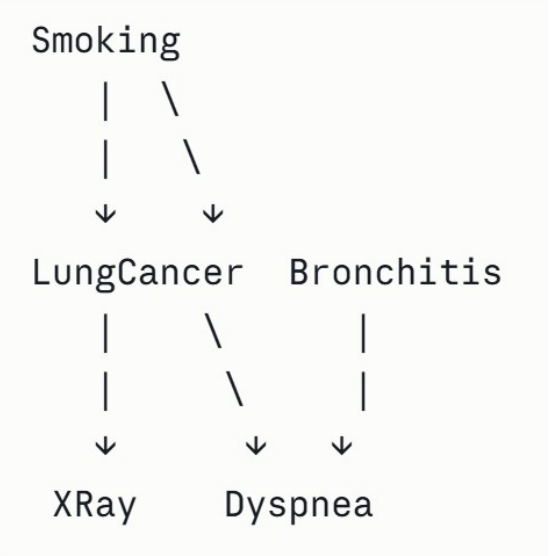


Example 2: Find the join factorization

$$P(S, L, B, X, D) = P(S, L, B, X, D) = P(S) \times P(L|S) \times P(B|S) \times P(X|L) \times P(D|L,B)$$

Where:

S=Smoking,
L=LungCancer,
B=Bronchitis,
X=XRay,
D=Dyspnea.



Example2: Given CPT

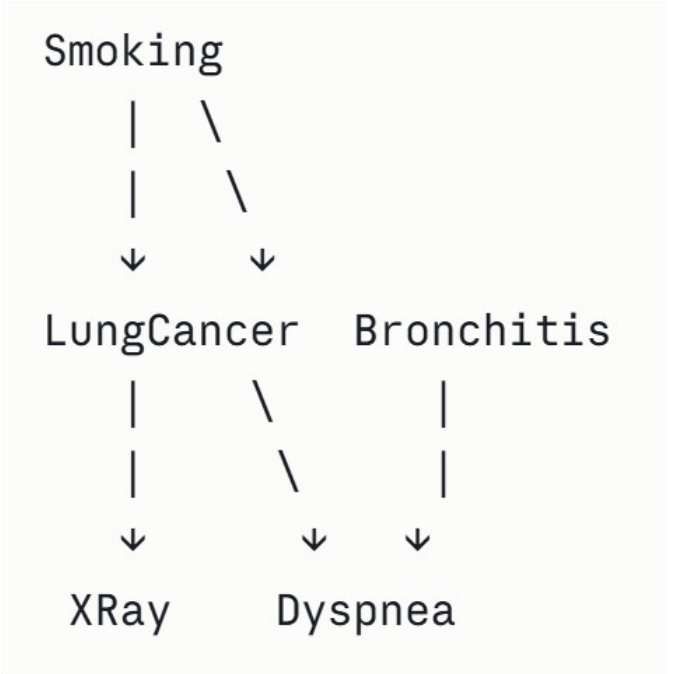
Smoking	P	P(S)
Yes	0.30	
No	0.70	

Smoking	P(Yes)	P(No)	P(L|S)
Yes	0.10	0.90	
No	0.01	0.99	

Smoking	P(Yes)	P(No)	P(B|S)
Yes	0.60	0.40	
No	0.30	0.70	

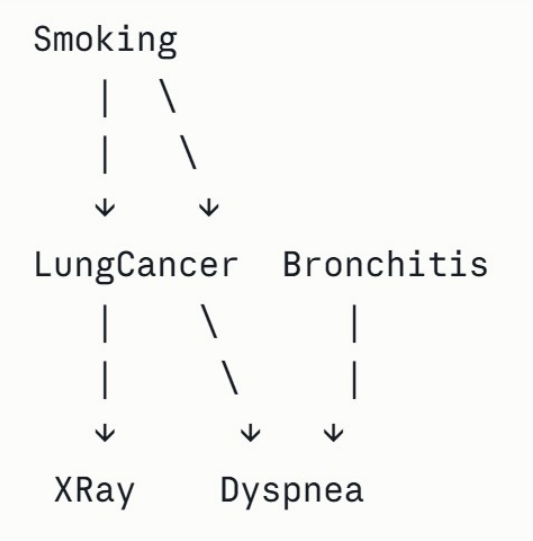
LungCancer	P(Positive)	P(Negative)	P(X | L)
Yes	0.90	0.10	
No	0.05	0.95	

LungCancer	Bronchitis	P(Yes)	P(No)	P(D|L,B)
Yes	Yes	0.98	0.02	
Yes	No	0.90	0.10	
No	Yes	0.70	0.30	
No	No	0.05	0.95	



Inference

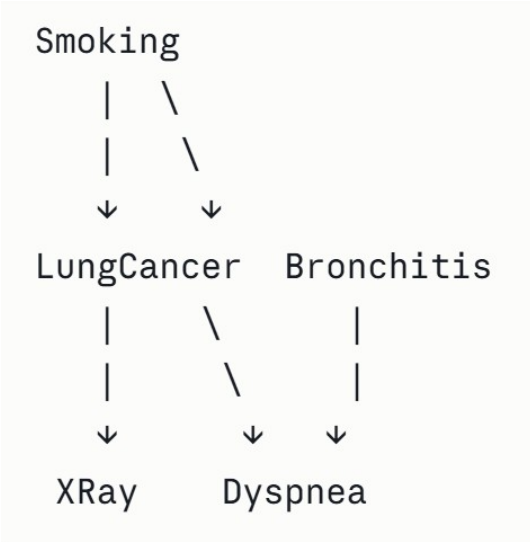
P(Dyspnea | LungCancer=Yes, Bronchitis=No) = ?



Inference

P(Dyspnea | LungCancer=Yes, Bronchitis=No) = ?

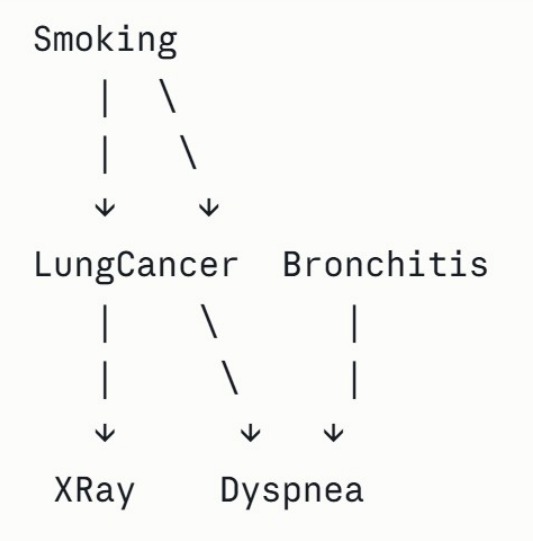
LungCancer	Bronchitis	P(Yes)	P(No) P(D| L,B)
Yes	Yes	0.98	0.02
Yes	No	0.90	0.10
No	Yes	0.70	0.30
No	No	0.05	0.95



inference

What is the probability of picking a random person who has LungCancer?

P(L) = ?



inference

What is the probability of picking a random person who has LungCancer?

P(L) = P(L) = Σ over S: P(L | S) × P(S), meaning

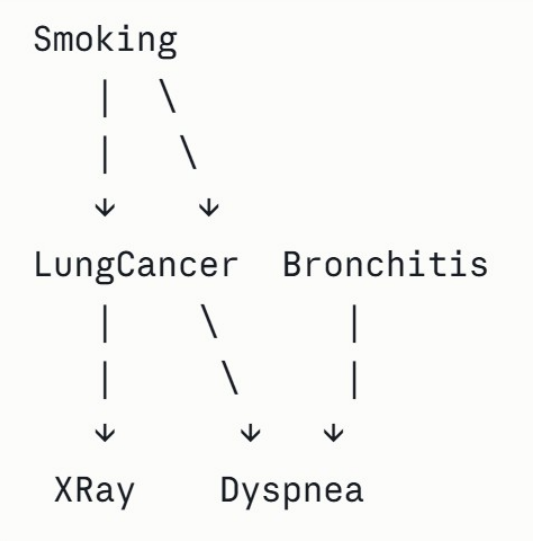
P(L)=P(L) = P(L | S=Yes) × P(S=Yes) + P(L | S=No) × P(S=No)

P(L=Yes) = P(L=Yes | S=Yes) × P(S=Yes) + P(L=Yes | S=No) × P(S

= 0.10 × 0.30 + 0.01 × 0.70

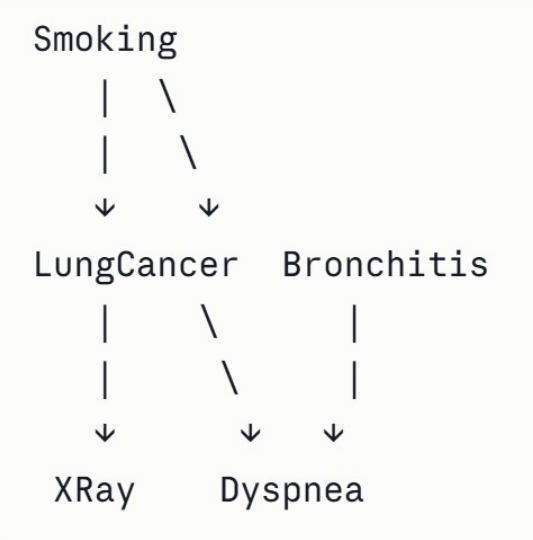
= 0.030 + 0.007

= 0.037



Inference

- What is the probability of picking a random person with not Bronchitis?
- P(B) = Σ over S: P(B | S) × P(S) =



inference

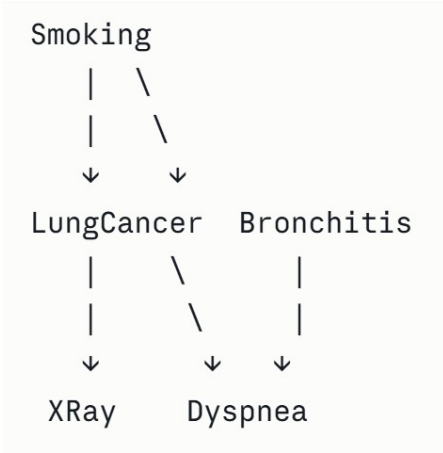
$$P(B) = P(B) = P(B \mid S=Yes) \times P(S=Yes) + P(B \mid S=No) \times P(S=No)$$

$$P(B=Yes) = P(B=Yes \mid S=Yes) \times P(S=Yes) + P(B=Yes \mid S=No) \times P(S=No)$$

$$= 0.60 \times 0.30 + 0.30 \times 0.70$$

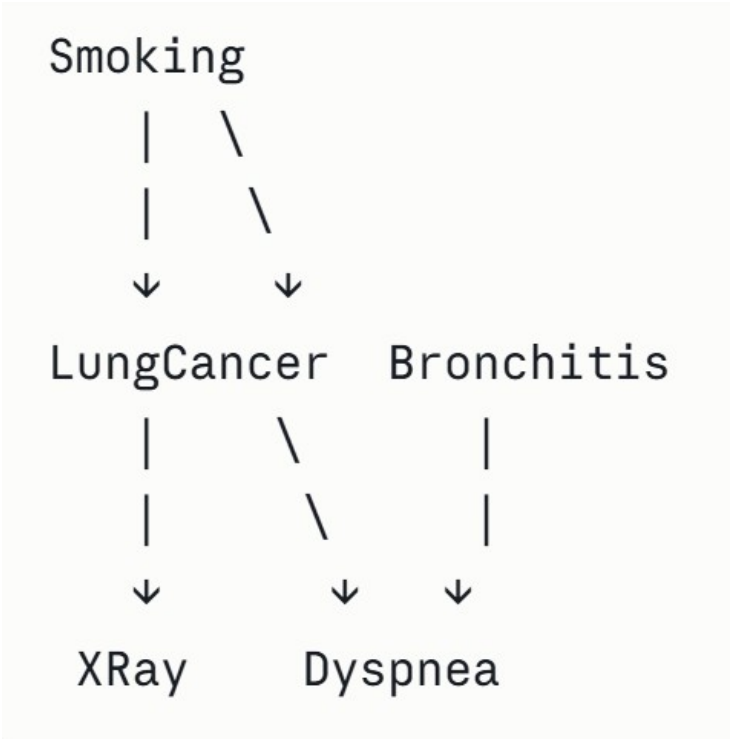
$$= 0.180 + 0.210$$

$$= 0.390$$



inference

What is the probability P(XRAY = positive

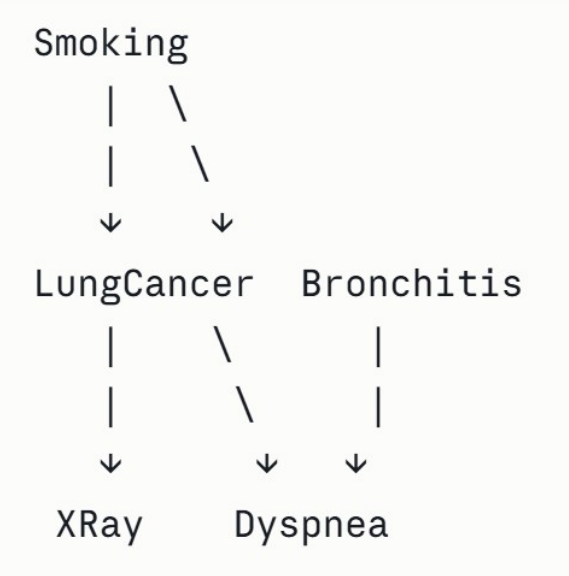


inference

What is the probability P(XRAY = positive)?

P(XRAY = positive)= Σ over L:P(Xray | L)P(L)

P(X=Positive) = P(X=Positive | L=Yes) × P(L=Yes) + P(X=Positive | L=N



One more inference

$P(L \mid S, B, X, D) = ?$

ONE more inference

- P(L | S, B, X, D) = P(L,S,B,X,D)/P(S,B,X,D)

=